Marginal structure models

usually estimated with inverse-probability-of-treatment weighting (IPTW) but not limited to IPTW

Sukie Xiuqi Yang
11/16/2022

Outline

- Marginal structural models (MSMs): What and why?
- When would you apply them?
  - Examples
- Limitations and some remedies
- New developments – different ways to construct weights
  - Covariate balancing propensity scores (CBPS) (Imai and Ratkovic 2014, 2015)
  - Residual balancing weights (Zhou and Wodtke 2020; Baum and Zhou Forthcoming)
- How do you execute? (example codes)

What are marginal structural models?

Multi-step estimation process:

- Separates confounder control from model estimation for effect of interest

Estimation process involves:

- Calculating weights
- Running a model (or other estimation procedure) using the weights

Mostly commonly applied for:

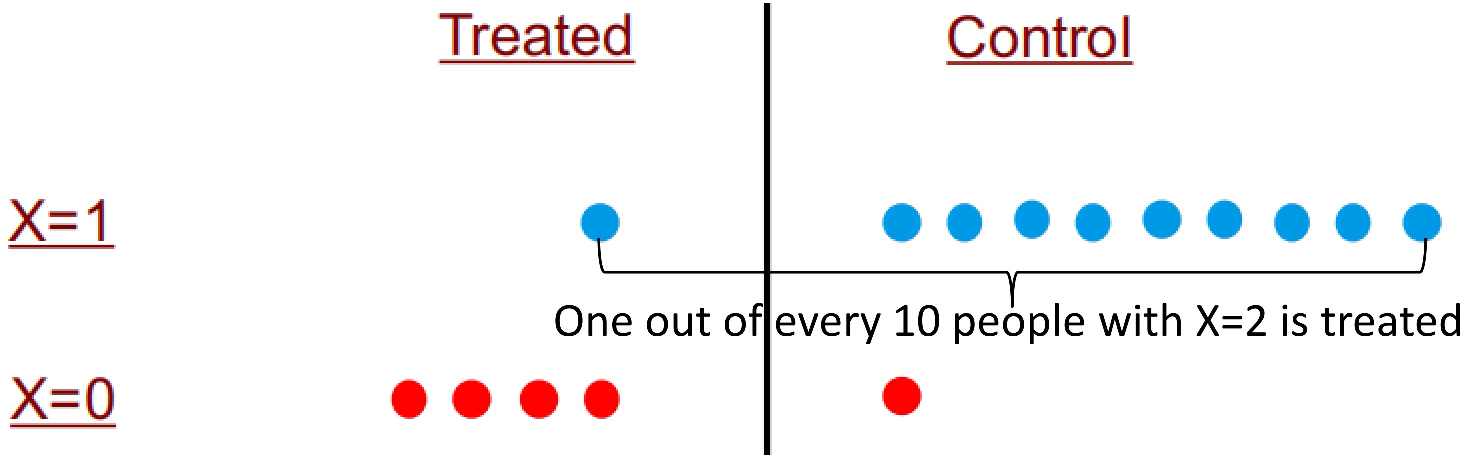
- Casual inference on observational data
- Causal mediation
- Controlling time-varying confounding

What are marginal structural models?

Why these names?

- "Marginal" in a sense that this class of models usually do not condition on other variables, but rely on the marginal distribution of the exposure while balancing confounders over the level of exposure.
- "Structural" is the econometric term for "causal"
- Inverse-probability-of-treatment weighting (IPTW) vs. Propensity scores
- History about the models (Xi might introduce it)

The intuition behind IPTW



Suppose that P(A=1|X=1)=0.1.

- Among people with X=1, only 10% will receive the treatment.
- i.e., the value of the propensity score for people with X=1 is 0.1.

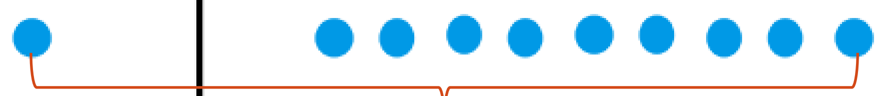
Suppose that P(A=1|X=0)=0.8.

- Among people with X=0, 80% will receive treatment.
- i.e., the value of the propensity score for people with X=0 is 0.8

The intuition behind IPTW

Treated | Control

X=1



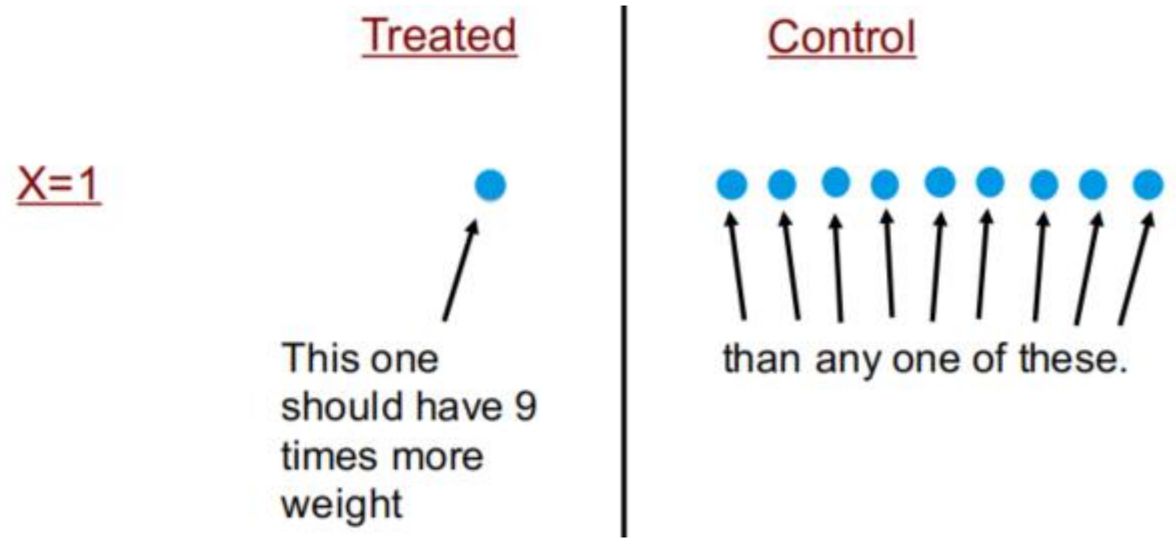
One out of every 10 people with X=2 is treated

Propensity score matching:

1 person in the treated group

Counts the same as 9 people from the control group

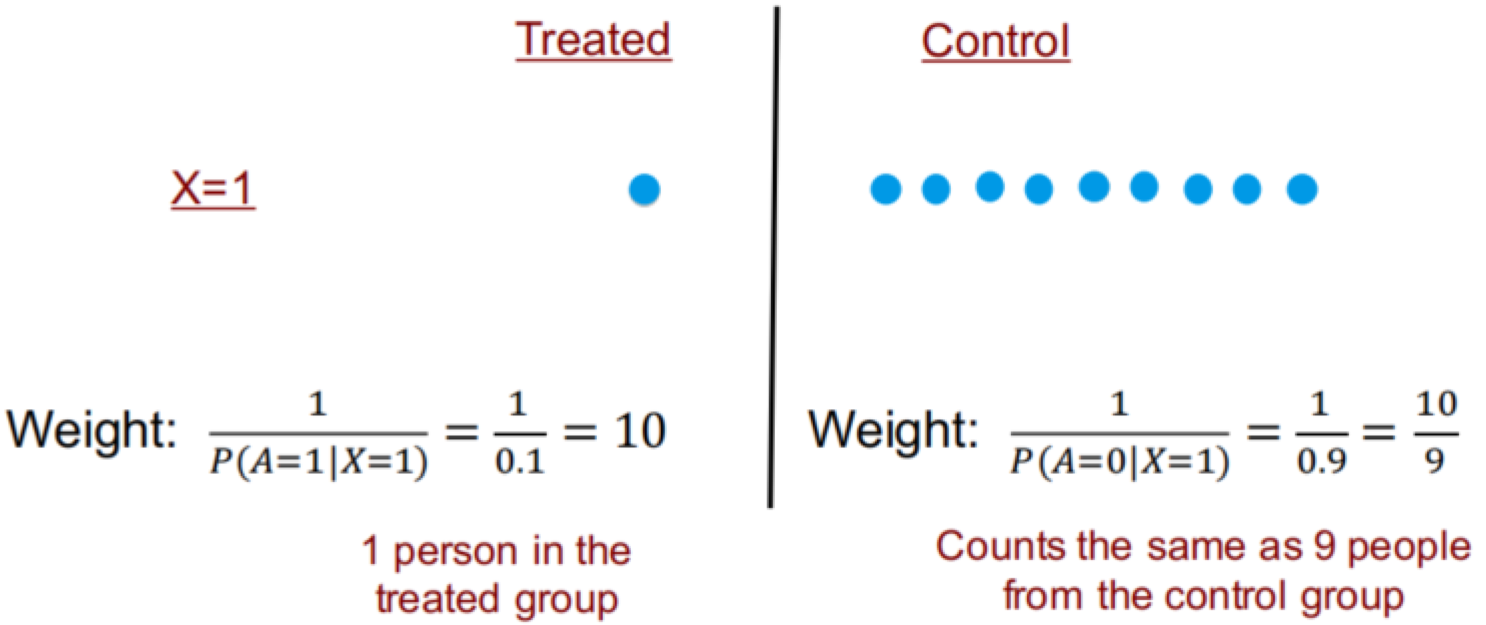
The intuition behind IPTW



Weighting: Rather than match, we could use all of the data, but down weight some and up weight others. This is accomplished by weighting by the inverse of the probability of treatment received.

For treated subjects weight by the inverse of P(A=1|X). For control subjects weight by the inverse of P(A=0|X), thus different from the propensity score.

The intuition behind IPTW



Weighting: Rather than match, we could use all of the data, but down weight some and up weight others. This is accomplished by weighting by the inverse of the probability of treatment received.

The intuition behind IPTW: survey sampling

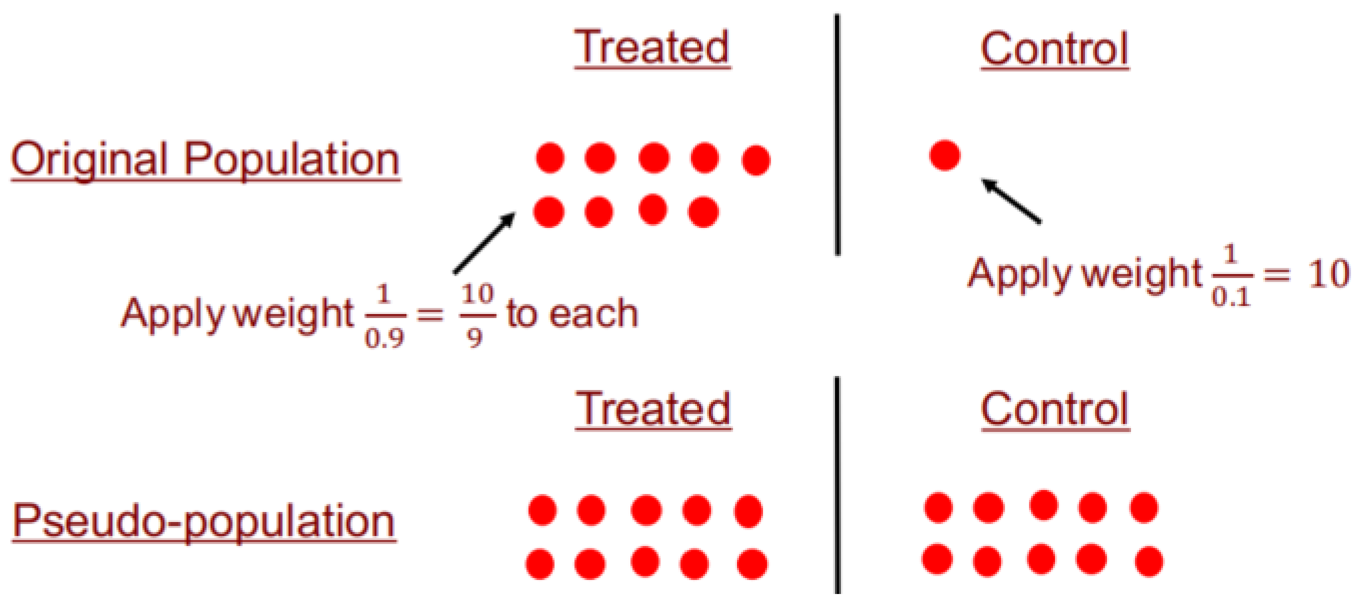
In surveys it is common to oversample some groups relative to the population.

- Oversample a minority group
- Oversample older adults
- Oversample obese individuals

To estimate the population mean, can weight the data to account for the oversample.

The intuition behind IPTW: Pseudo-population

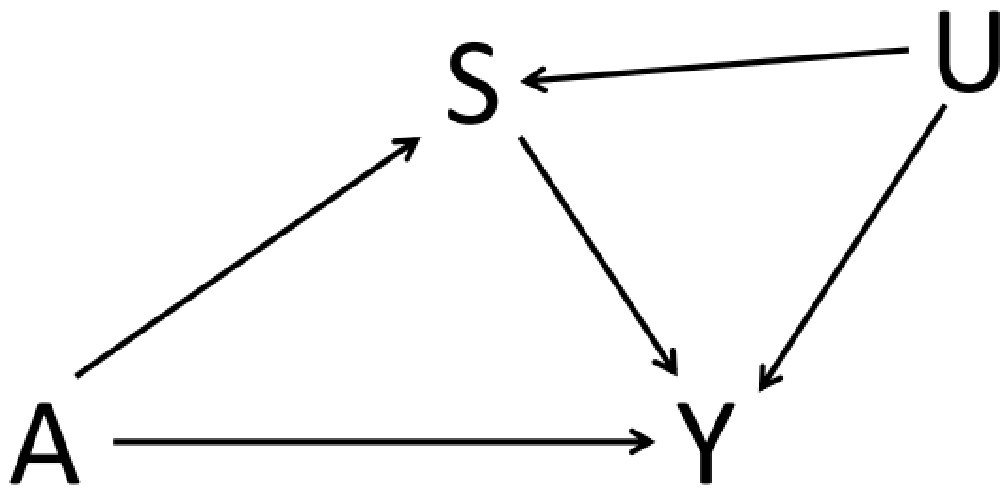
- Suppose P(A=1|X)=0.9



- In the original population, some people were more likely to get treated than others, based on their X's.
- In the pseudo-population, everyone is equally likely to be treated, regardless of their X values.

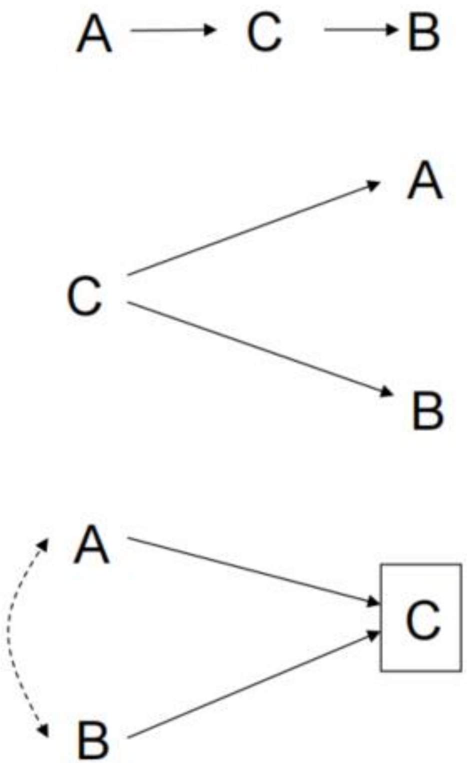
When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

Assuming A is randomized, S is the mediating variable, potential outcomes can be defined as $Y^{a,s}$, U are the unmeasured confounders.



Quick recap on the DAGs and the ignorability assumptions in causal inference

Three Sources of Association Between Two Variables A & B



(1) Direct and indirect causation

$A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B|C$

(2) Common cause confounding

$A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B|C$

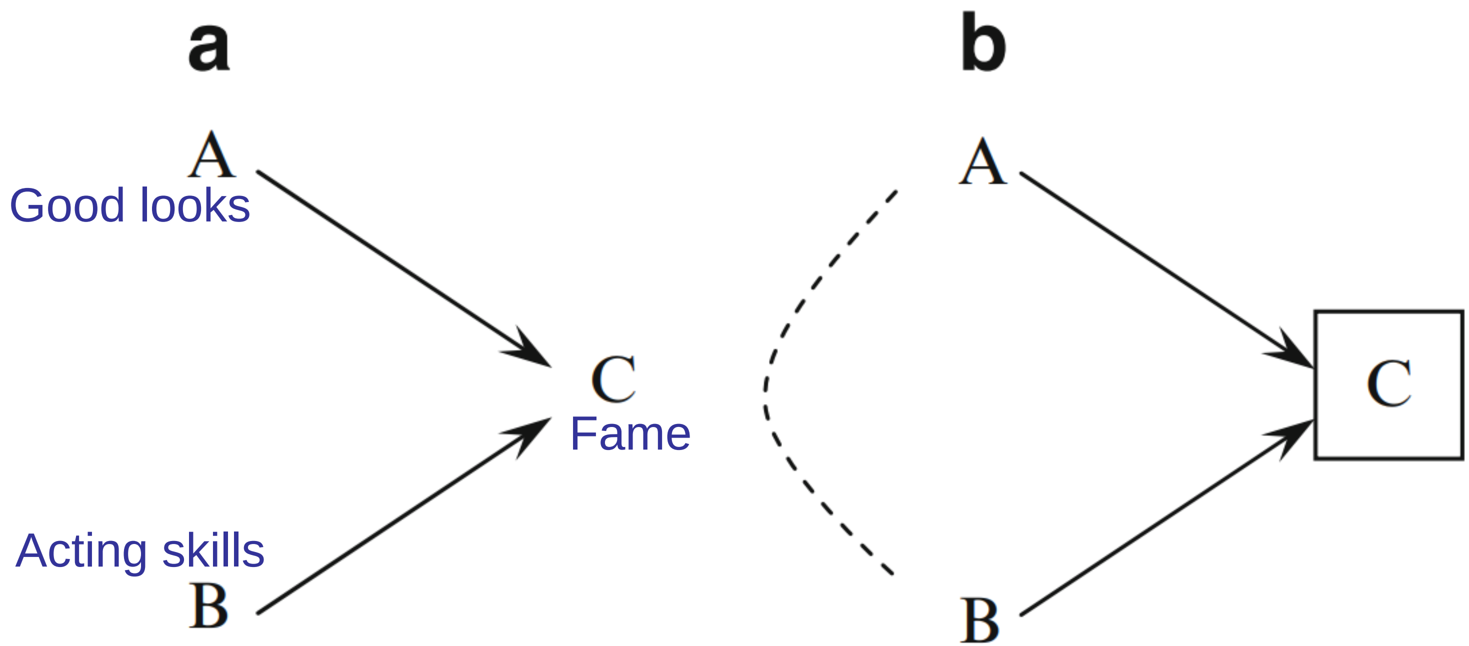
(3) Conditioning on a common effect ("collider"): Selection

$A \perp\!\!\!\perp B$ and $A \not\perp\!\!\!\perp B|C$

←-------→ : non-causal (spurious) association. ☐ : conditioning.

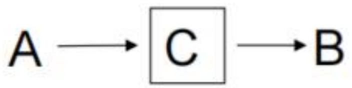
Quick recap on the DAGs and the ignorability assumptions in causal inference

Three Sources of Association Between Two Variables A & B

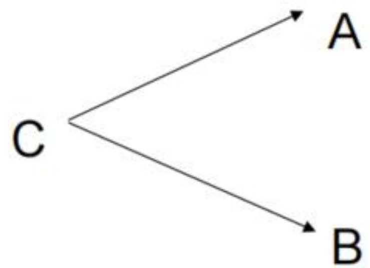


Quick recap on the DAGs and the ignorability assumptions in causal inference

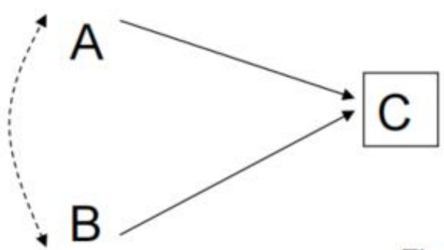
Three Sources of Association Between Two Variables A & B



Overcontrol: intercepting the causal pathway



Confounding bias: failure to condition on a common cause



Endogenous selection bias: mistaken conditioning on a common effect.

Elwert@wisc.edu. Version 5/2013

All three constitute analytic mistakes (Elwert 2013).

Quick recap on the DAGs and the ignorability assumptions in causal inference

Three blocking criteria (key!!)

1. Conditioning on a non-collider blocks a path
2. Conditioning on a collider, or a descendent of a collider, unblocks a path
3. Not conditioning on a collider leaves a path “naturally” blocked.

<u>The adjustment criterion reveals which variables give (conditional) ignorability.</u>

Probabilistic Implications

Two nodes, A and B, are d-separated by a set of nodes C iff it *blocks every path* from A to B.

- Then: $A \perp\!\!\!\perp B | C$

Quick recap on the DAGs and the ignorability assumptions in causal inference

- "Eliminate backdoor paths between treatment (A) to Y" in the DAG
  = d-separate/block every path between A and Y that contain an arrow into A while not conditioning on descendant of A (Pearl 1988)
- "Conditional exchangeability is often referred as 'weak ignorability' or 'ignorable treatment assignment' in statistics (Rosenbaum and Rubin, 1983), 'selection on observables' in the social sciences (Barnow et al., 1980), and 'no omitted variable bias' or 'exogeneity' in econometrics (Imbens, 2004)."

-- Hernán MA, Robins JM (2020). *Causal Inference: What If*.

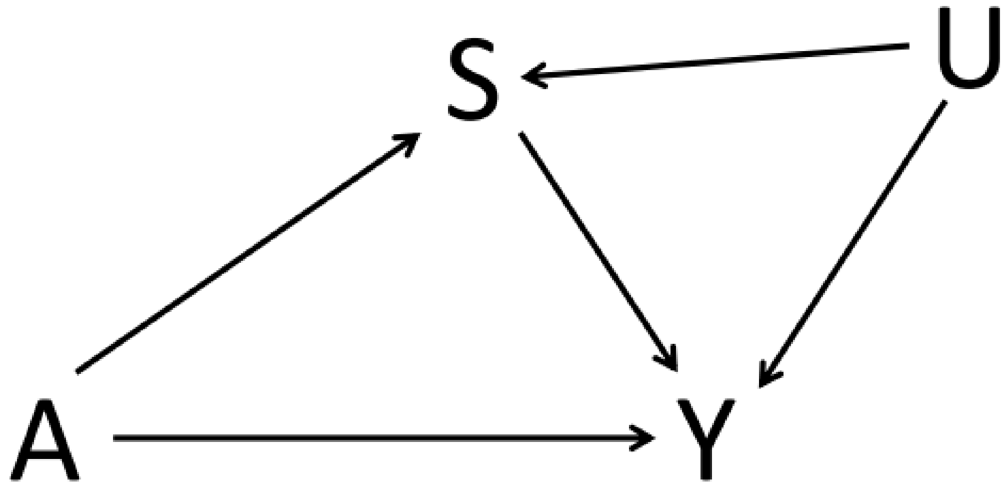
Quick recap on the DAGs and the ignorability assumptions in causal inference

- Eliminate backdoor paths between treatment (A) to Y" in the DAG = d-separate/block every path between A and Y that contain an arrow into A while not conditioning on descendant of A (Pearl 1988)
- "Conditional exchangeability is often referred as 'weak ignorability' or 'ignorable treatment assignment' in statistics (Rosenbaum and Rubin, 1983), 'selection on observables' in the social sciences (Barnow et al., 1980), and 'no omitted variable bias' or 'exogeneity' in econometrics (Imbens, 2004)."

-- Hernán MA, Robins JM (2020). *Causal Inference: What If*.

When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

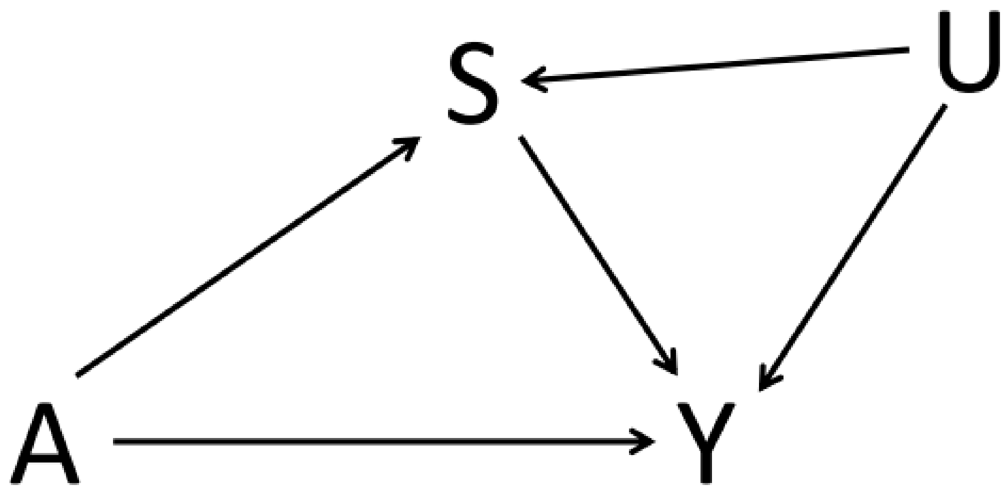
Assuming A is randomized, S is the mediating variable, potential outcomes can be defined as $Y^{a,s}$, U are the unmeasured confounders.



When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

Assuming *A* is randomized, *S* is the mediating variable, potential outcomes can be defined as $Y^{a,s}$, *U* are the unmeasured confounders.

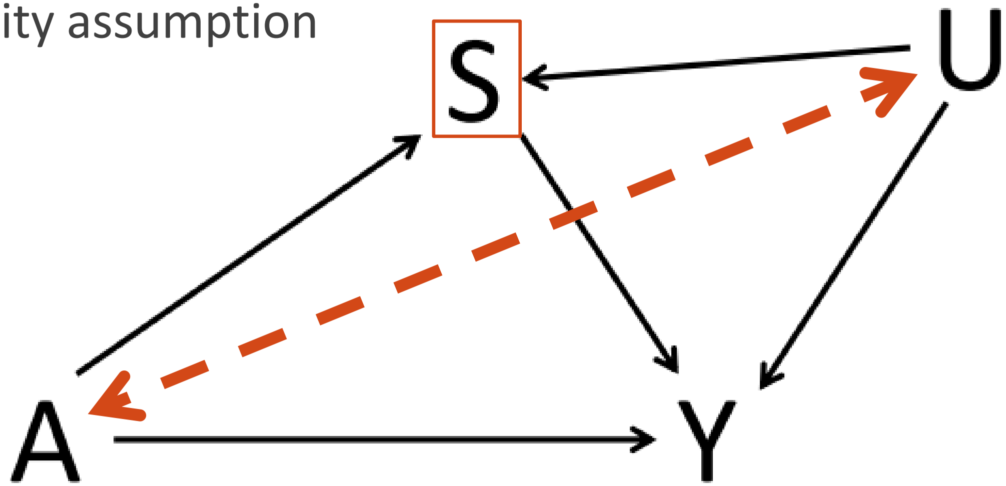
Fit a model of *Y* on *A* adjusting for *S*, i.e., $E(Y\ |A,S) = \beta_0 + \beta_1 A + \beta_2 S$, β_1 is biased due to the collider *S*.



When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

Assuming *A* is randomized, *S* is the mediating variable, potential outcomes can be defined as $Y^{a,s}$, *U* are the unmeasured confounders.

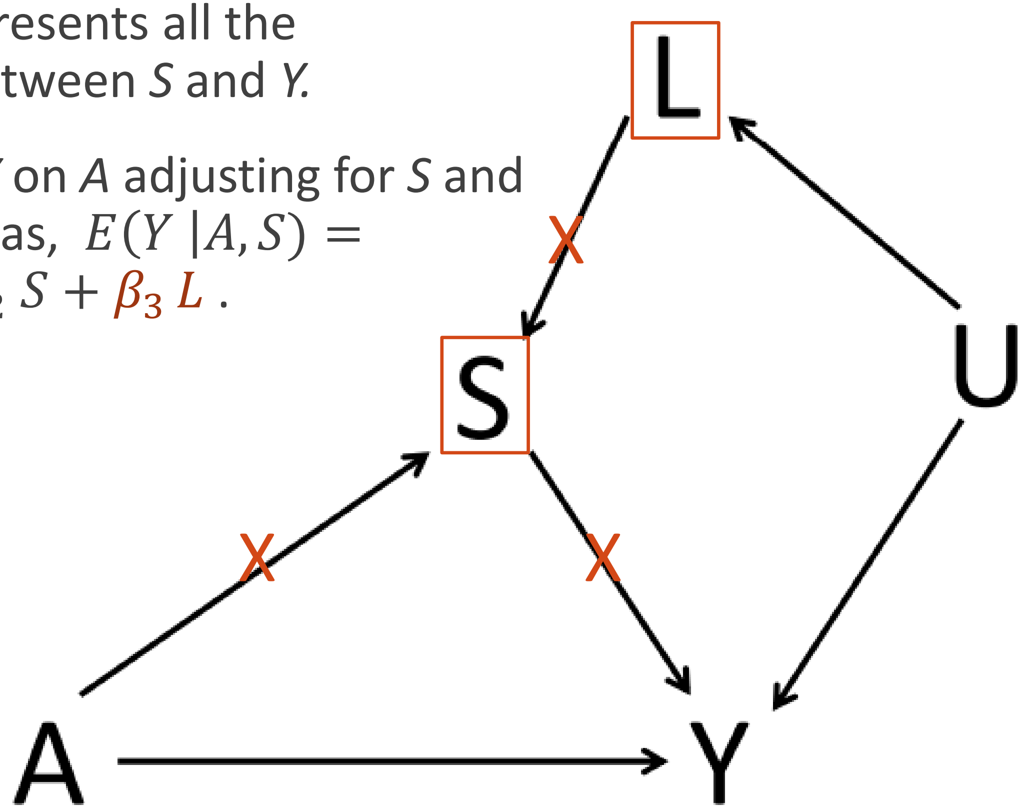
Fit a model of *Y* on *A* adjusting for *S*, i.e., $E(Y\mid A,S) = \beta_0 + \beta_1 A + \beta_2 S$, β_1 is biased due to the collider *S*. This opens up backdoor path between A and *A* -> *U* -> *Y*, which means that the ignobility assumption Is violated.



When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

Assuming *L* represents all the confounding between *S* and *Y*.

Fit a model of *Y* on *A* adjusting for *S* and *L* remove the bias, $E(Y \mid A, S) = \beta_0 + \beta_1 A + \beta_2 S + \beta_3 L$.

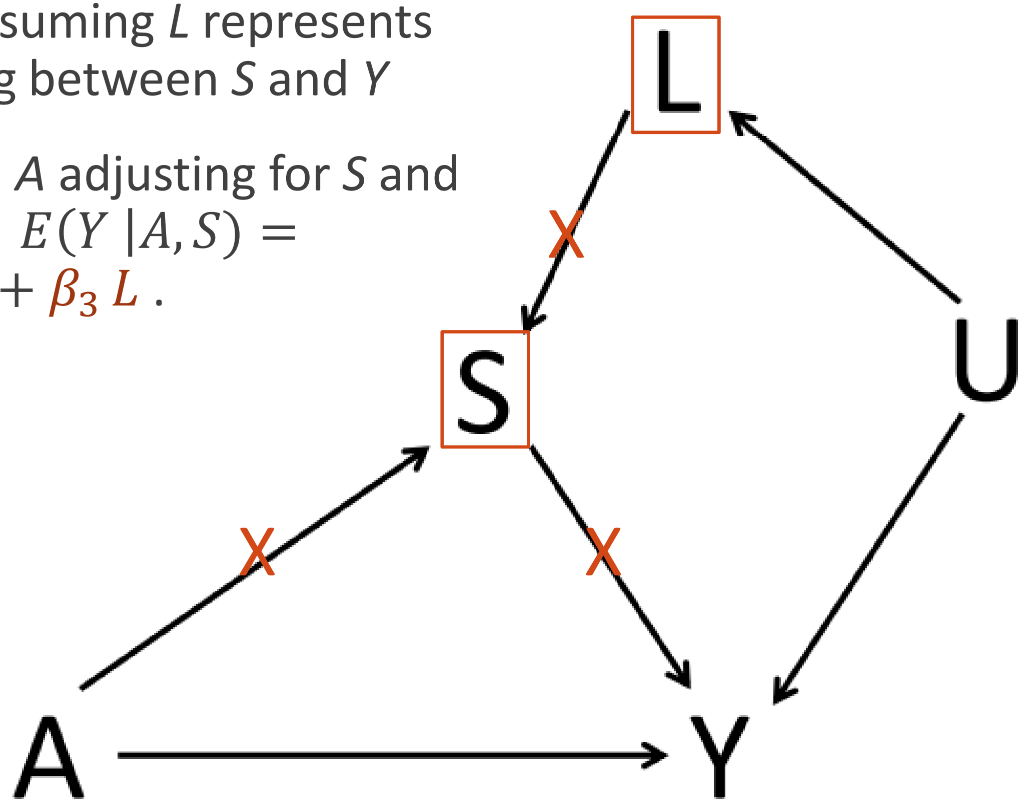


When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

In this example, assuming *L* represents all the confounding between *S* and *Y*

Fit a model of *Y* on *A* adjusting for *S* and *L* remove the bias, $E(Y \mid A, S) = \beta_0 + \beta_1 A + \beta_2 S + \beta_3 L$.

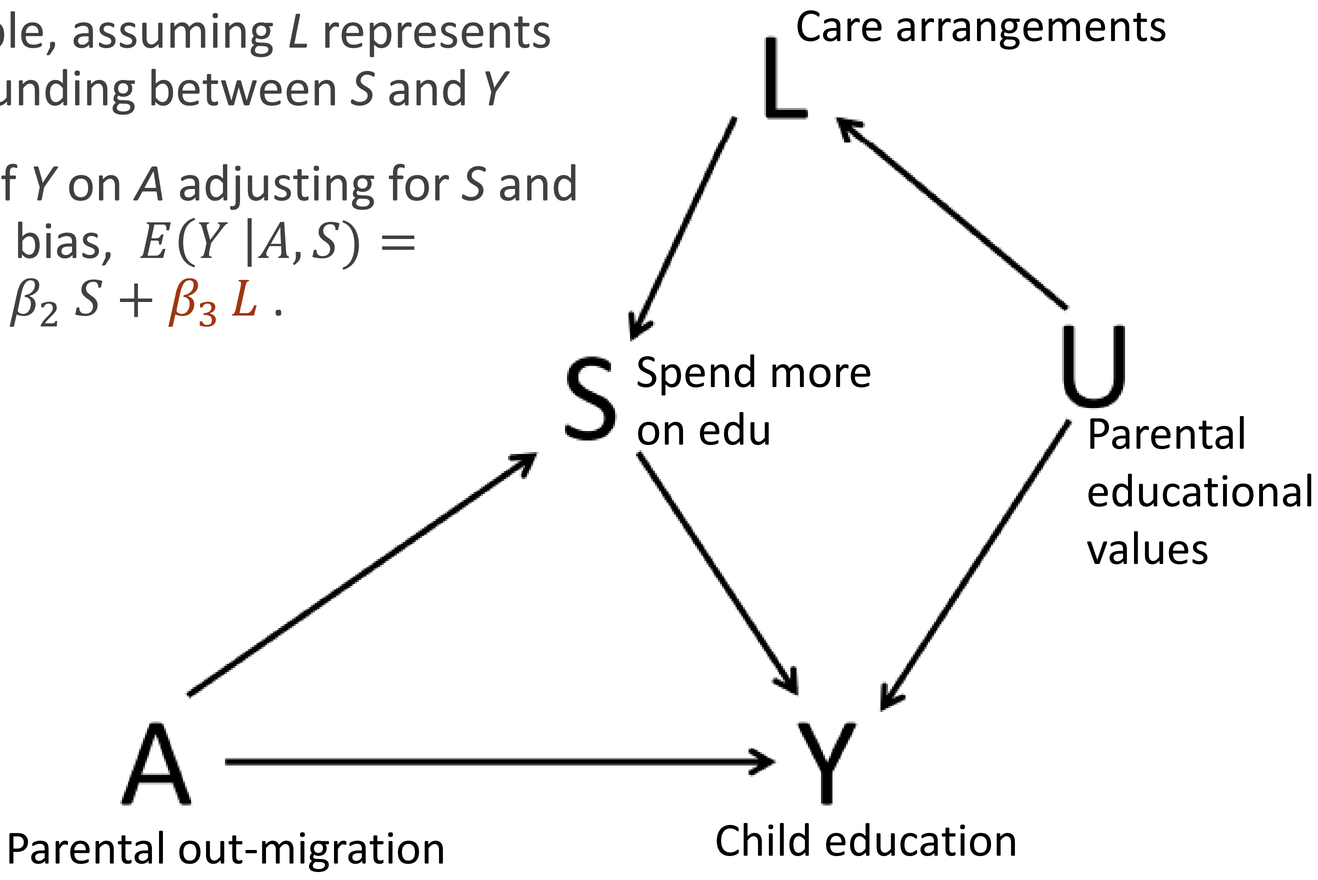
β_1 is unbiased.



When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

In this example, assuming L represents all the confounding between S and Y

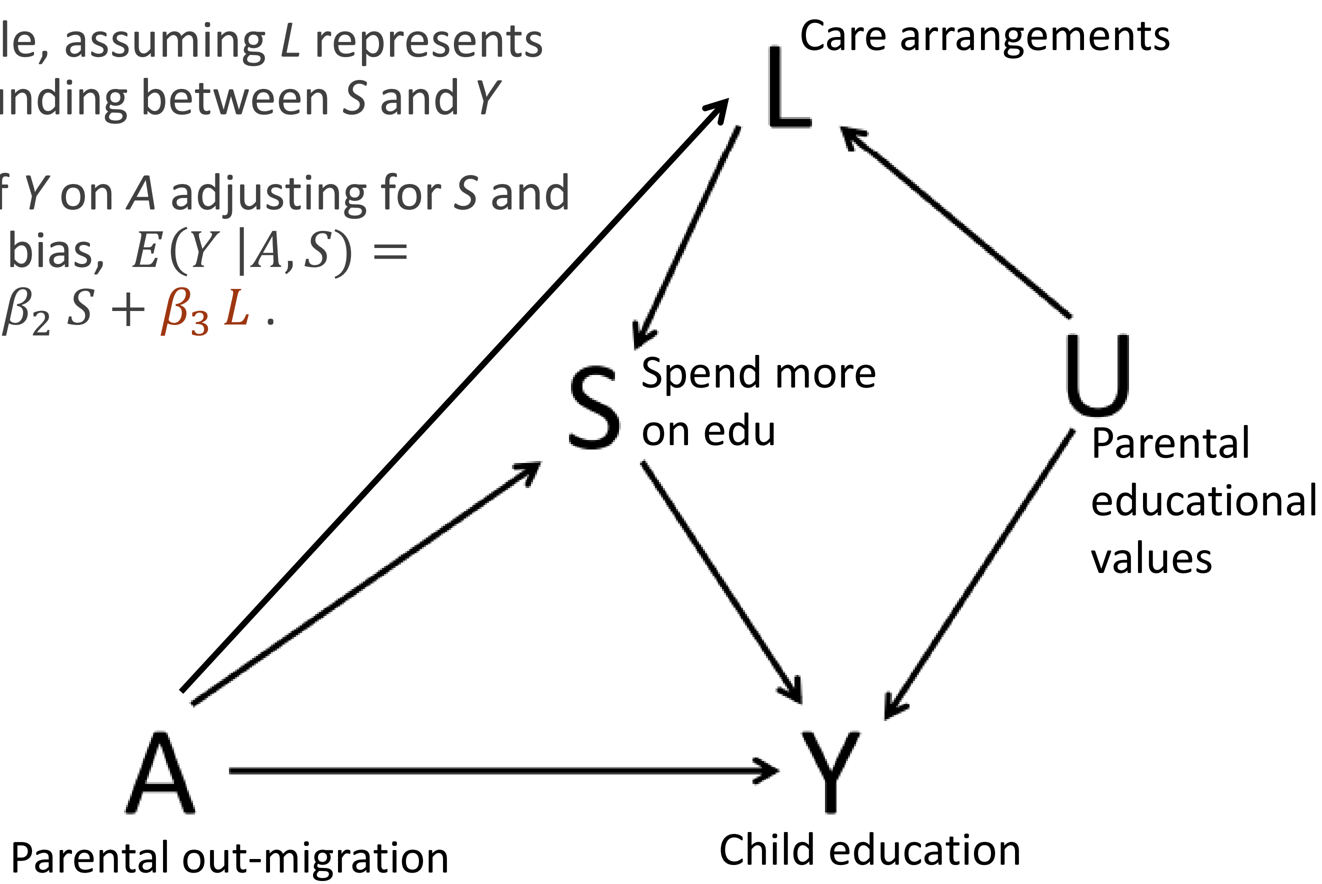
Fit a model of Y on A adjusting for S and L remove the bias, $E(Y\,|A,S) = \beta_0 + \beta_1\,A + \beta_2\,S + \beta_3\,L$.



When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

In this example, assuming *L* represents all the confounding between *S* and *Y*

Fit a model of *Y* on *A* adjusting for *S* and *L* remove the bias, $E(Y\,|A,S) = \beta_0 + \beta_1\,A + \beta_2\,S + \beta_3\,L$.

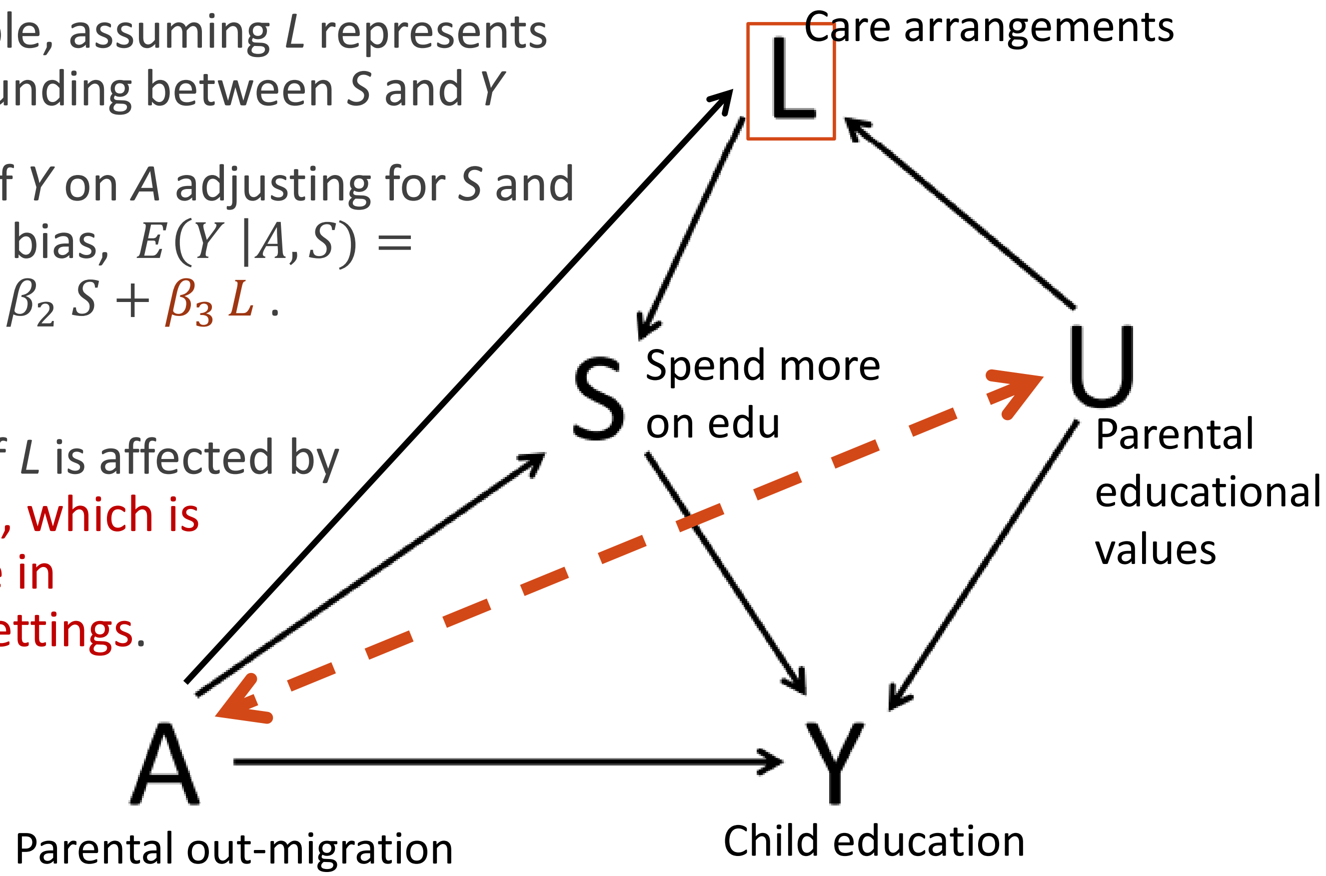


When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

In this example, assuming *L* represents all the confounding between *S* and *Y*

Fit a model of *Y* on *A* adjusting for *S* and *L* remove the bias, $E(Y \mid A, S) = \beta_0 + \beta_1 A + \beta_2 S + \beta_3 L$.

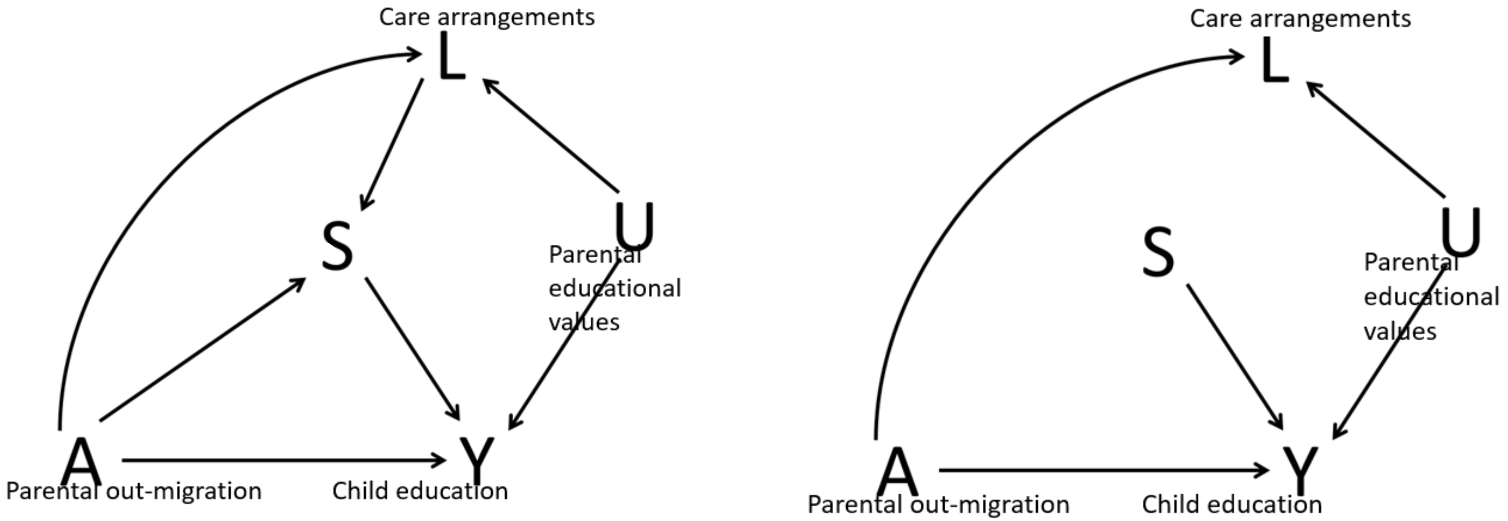
β_1 is biased if *L* is affected by the treatment, which is often the case in longitudinal settings.



When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

Robins (1999) proposed to weight the data by

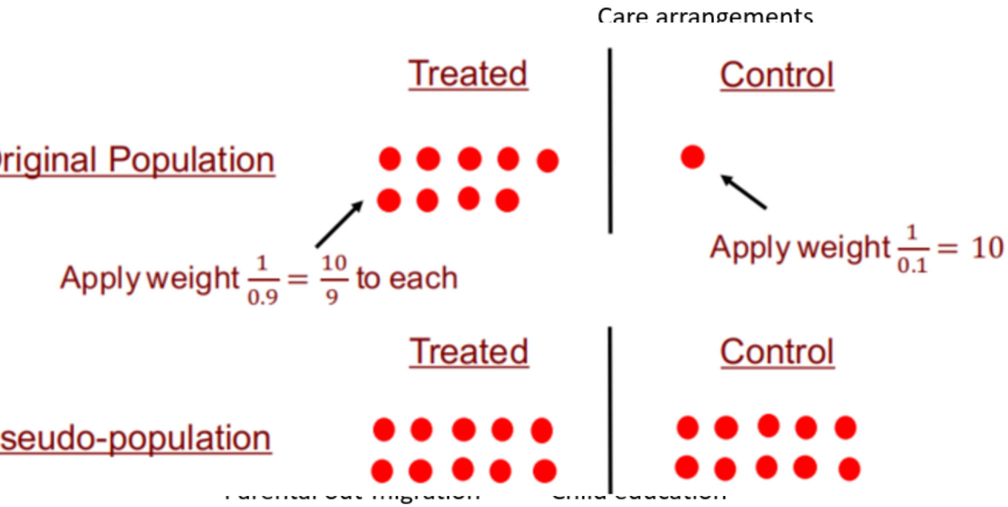
$$\frac{1}{f(S|A,L)}$$

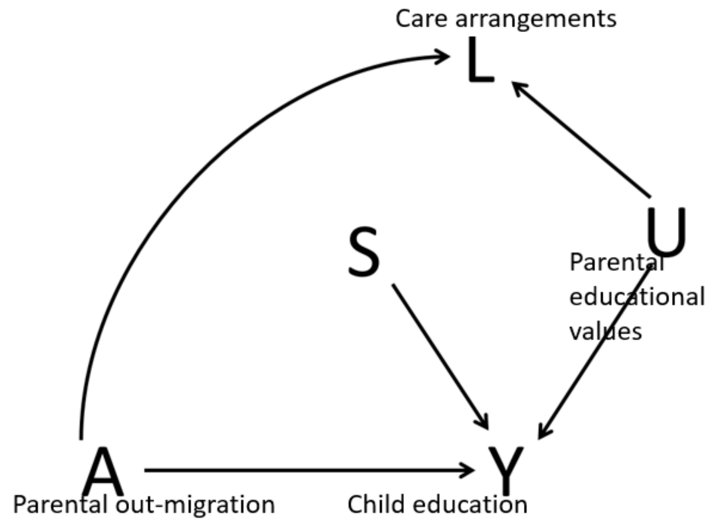


When would we use MSMs with IPTW? – First, consider a cross-sectional mediation example

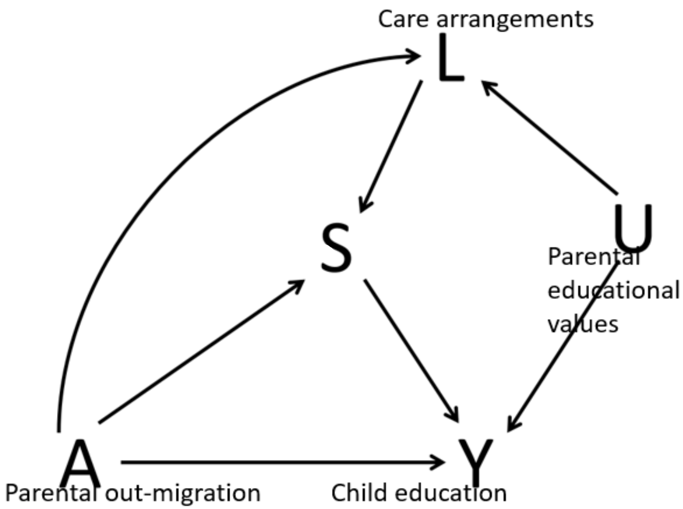
Robins (1999) proposed to weight the data by

$$\frac{1}{f(S|A,L)}$$

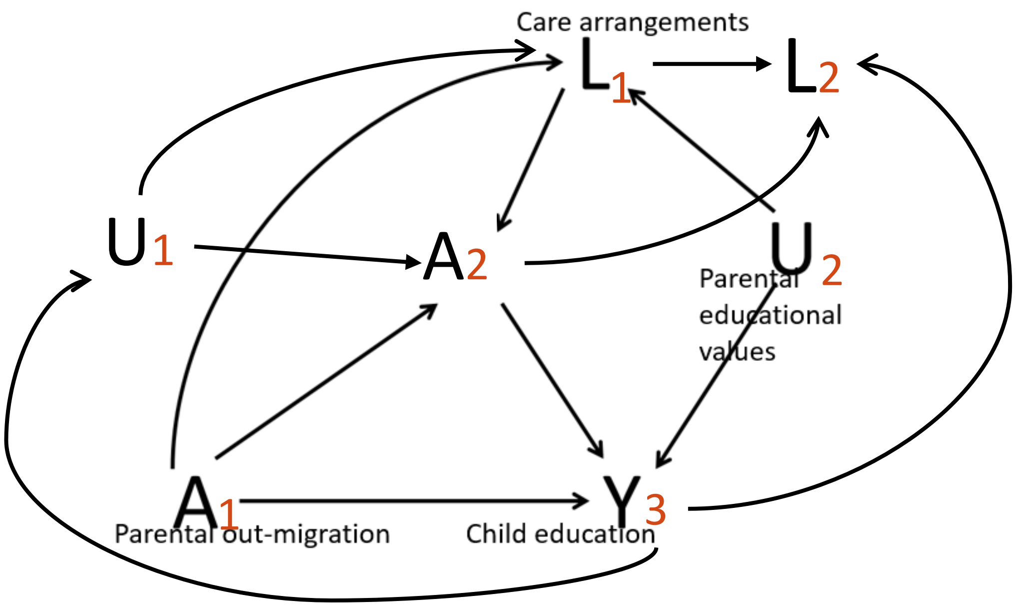




When would we use MSMs with IPTW? – Let’s extend the cross-sectional example



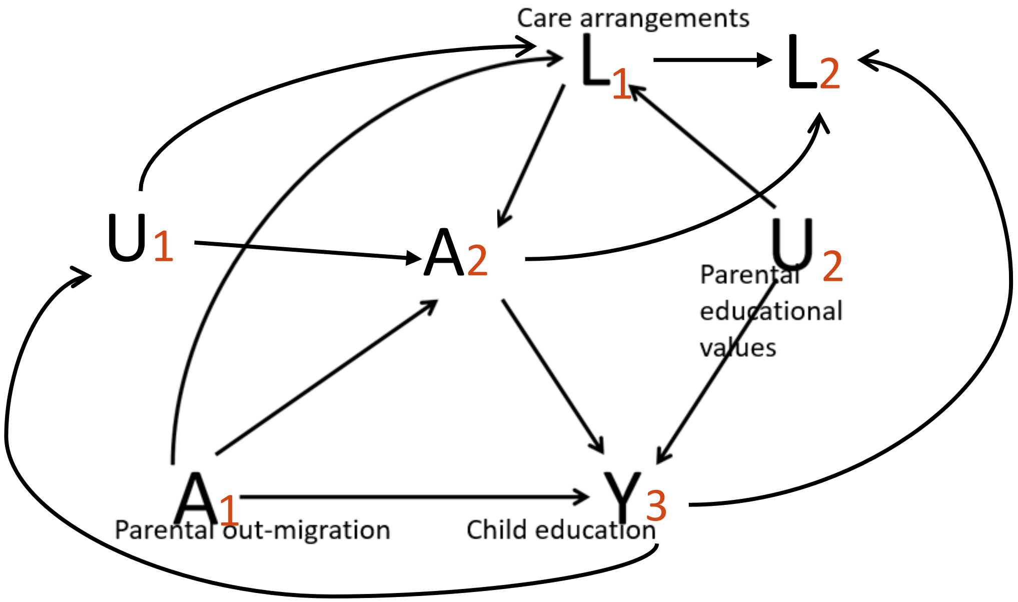
When would we use MSMs with IPTW? – Let’s extend the cross-sectional example



Subscripts denote wave or occasions

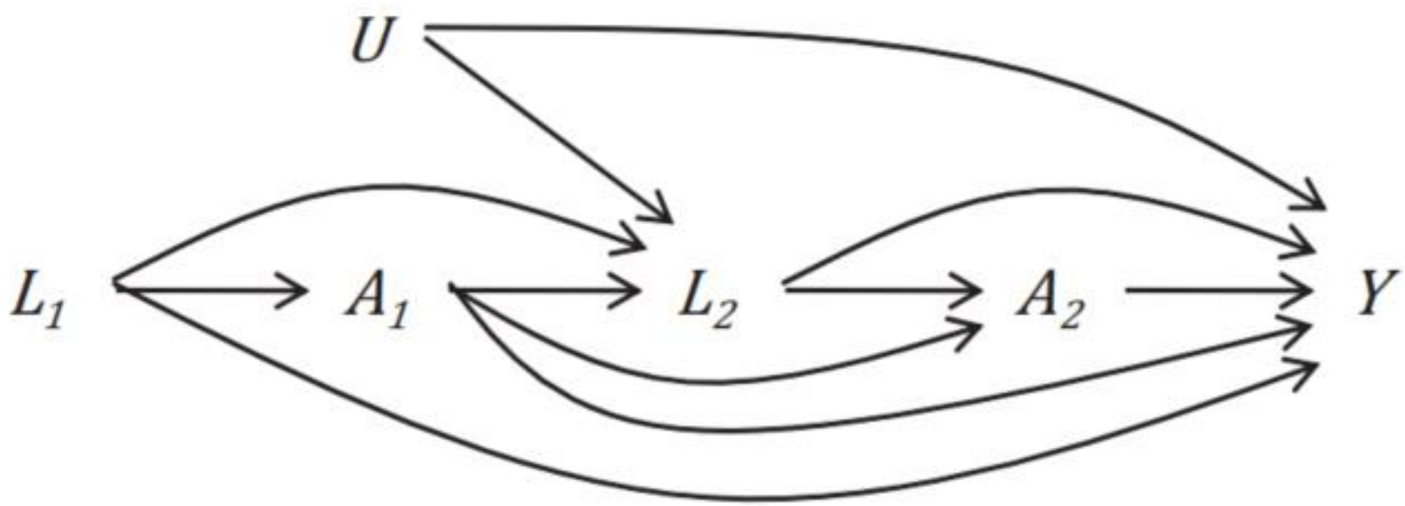
When would we use MSMs with IPTW? – Let’s extend the cross-sectional example

Looks horrible; let’s tidy it up!

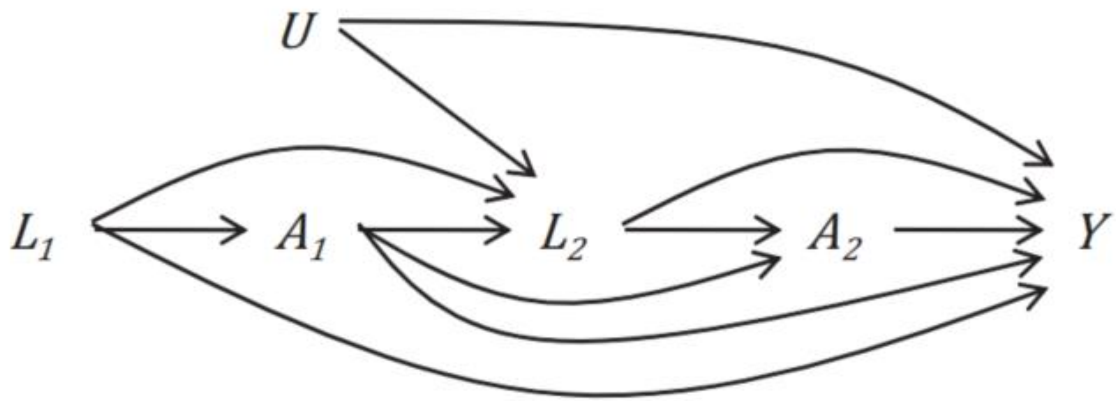


Subscripts denote wave or occasions

When would we use MSMs with IPTW? – Let’s extend the cross-sectional example



When would we use MSMs with IPTW? – Causal Effects of Time-Varying Treatment



- Consider joint effects of multiple treatments $A_1, \cdots, A_m$
- Consider what would happen if received treatment levels $a_1, \cdots, a_m$
- L_k denotes covariate levels at time k
- $L_k \equiv \{L_0, L_1, \cdots, L_k\}$ denotes covariate history through k

Compare potential outcomes under different regimes Y^g, $Y^{g'}$, $g, g' \in G$

(Regime: a plan, analogous to protocol in clinical trial, which specifies what treatment a subject is to receive at any point in time)

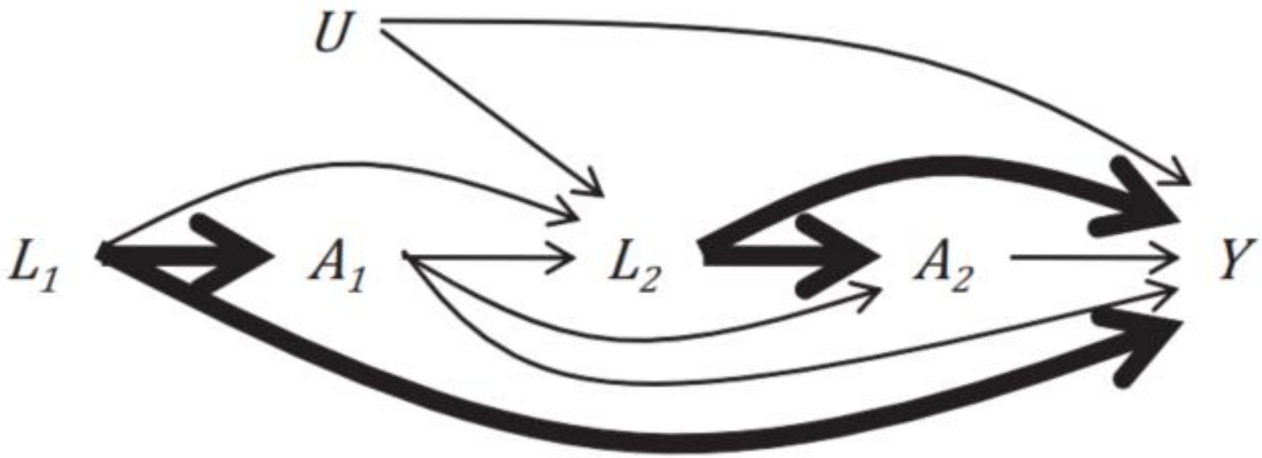
What levels of treatment are appropriate choices of comparison?

- What levels of treatment are appropriate choices of comparison?
  - Local comparison: $Y^{a_1,a_2,\cdots,a_m}$ to $Y^{a_1',a_2,\cdots,a_m}$ or $Y^{a_1,a_2',\cdots,a_m}$
    - Treatments received differ only in 1 element; at 1 point in time
    - e.g., effect of 1 day of AZT
  - Global comparison: $Y^{a_1,a_2,\cdots,a_m}$ to $Y^{a_1',a_2',\cdots,a_m'}$
    - Treatments received differ in many elements; at many points in time
    - e.g., effect of several days of AZT

How do MSMs help address time-varying confounding?

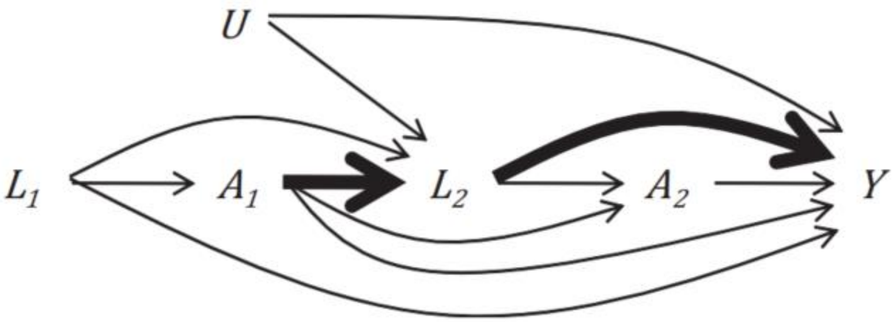
- They do not fix confounders as a method of adjustment (like in regression)
- Weighting produces the 'pseudo-population' in which all confounders (including those that vary with time) are balanced.

1. Not conditioning on L_k confounding

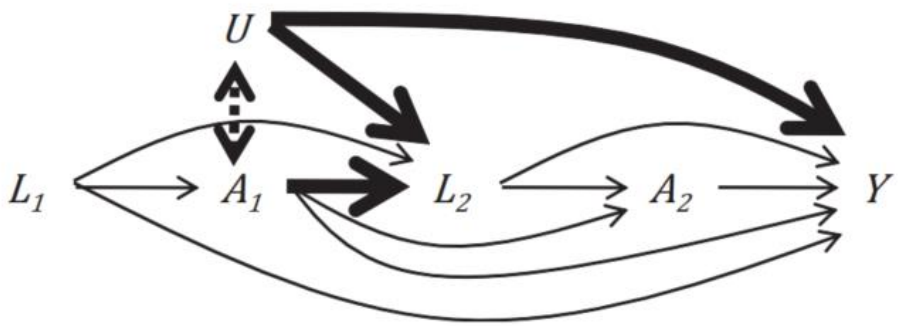


How do MSMs help address time-varying confounding?

2. Conditioning on L_k over-controls indirect pathways

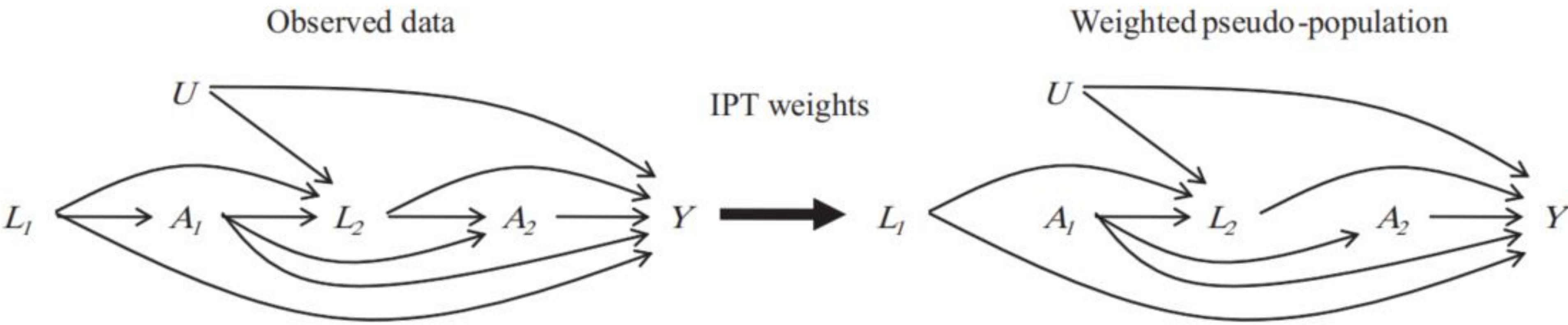


3. Conditioning on L_k incurs collider-stratification bias



How do MSMs help address time-varying confounding?

4. the Effect of Weighting by the Inverse Probability of Treatment (IPT)



Measured confounders are no longer confounders because there is no longer a relationship from L_k (time-varying confounder) to A_k (exposure), enabling an unbiased exposure estimate. It imitates sequentially randomized experiment.

IPTW Assumptions

Conditional ignorability/exchangability (for scalar treatment) or Sequential exchangeability (for time-varying treatment)

- Absence of unmeasured confounding
- Not directly testable; use theory and causal graphs/logic
- Sensitivity analyses can be used to quantify the impact of unmeasured confounding

Consistency (Sequential version for time-varying treatment)

- No misclassification of exposure
- Sensitivity analyses

Positivity -i.e. a non-zero (or 1) probability of receiving treatment

- Can't have perfect confounder combination to determine treatment or non-treatment

Correctly specified IPTW (from the model)

- Assumptions of the statistical model used to generate the IPTW are met

IPTW for time-varying treatment

- $$w_i = \prod_{t=1}^{T} \frac{1}{f(A_t = a_{i,t} | \overline{A}_{t-1} = \overline{a}_{i,t-1}, \overline{L}_t = \overline{l}_{i,t})},$$

covariate history $\bar{L}_t$ and treatment history $\bar{A}_{t-1}$ at all time t, $t = 1, \cdots, T$

- "stabilized" weight:

$$sw_i = \prod_{t=1}^{T} \frac{f(A_t = a_{i,t} | \overline{A}_{t-1} = \overline{a}_{i,t-1})}{f(A_t = a_{i,t} | \overline{A}_{t-1} = \overline{a}_{i,t-1}, \overline{L}_t = \overline{l}_{i,t})}$$

IPTW for time-varying treatment

- "stabilized" weight (cont.):

$$sw_i = \prod_{t=1}^{T} \frac{f(A_t = a_{i,t} | \overline{A}_{t-1} = \overline{a}_{i,t-1}, X = x)}{f(A_t = a_{i,t} | \overline{A}_{t-1} = \overline{a}_{i,t-1}, \overline{L}_t = \overline{l}_{i,t}, X = x)}$$

covariate history $\bar{L}_t$ and treatment history $\bar{A}_{t-1}$ at all time *t*, *t* = 1, · · · ,*T*.

X is a set of baseline or time-invariant confounders.

Notes: In such cases, these variables need to be included in the MSM to properly adjust for confounding, which is unproblematic because they cannot be affected by treatment.

Limitations of IPTW estimation

- IPTW estimation is relatively inefficient.
  - Remedy could be augmenting the IPTW estimator by additionally adjusting for confounders directly in the outcome regression also improves its efficiency (Robins et al. 1994).
- It is susceptible to finite-sample bias.
- IPTW estimation can be difficult to implement with continuous treatments (Zhou and Wodtke 2020; Wodtke 2018).
  - Residual balancing weights (Zhou and Wodtke 2020; Baum and Zhou Forthcoming)
- Cannot estimate effect modification beyond the baseline, time-invariant covariates.
  - Structural Nested Models (Wodtke, Elwert, and Harding 2016)

Example 1: Neighborhood effect (Wodtke 2013; Wodtke, Harding, and Elwert, 2011)

Research question: How do the duration and timing of exposure to neighborhood poverty impact the risk of adolescent parenthood?

- Data: PSID
- Outcome (*Y*): adolescent parenthood
- Time-varying exposure ($A_0, \ldots A_k$): Level of neighborhood poverty

In sum, a number of time-varying family characteristics—parental employment, income, and family structure, in particular—affect future neighborhood selection and are themselves affected by past neighborhood contexts. Because these factors also influence the risk of adolescent parenthood (Duncan et al. 1998; McLanahan and Percheski 2008), they are simultaneously confounders for the effect of future exposures and mediators for the effect of past exposures to neighborhood poverty. Time-varying

Example 1: Neighborhood effect (Wodtke 2013; Wodtke, Harding, and Elwert, 2011)

$$\text{logit}\left(P\left(Y_k(\overline{a}) = 1 \middle| k > 7, \overline{Y}_{k-1}(\overline{a}) = 0\right)\right) = \beta_0(k) + \beta_1\left(\frac{\sum_{t=1}^{k-1} I(a_t = 2)}{k-1}\right) + \beta_2\left(\frac{\sum_{t=1}^{k-1} I(a_t = 3)}{k-1}\right),$$

the proportion of time that subjects live in moderate- and high-poverty neighborhoods, respectively, from one wave post-baseline (i.e., age 5) through wave k – 1.
The author also looked at timing of exposure:

$$\text{logit}\left(P\left(Y_k(\overline{a}) = 1 \middle| k > 7, \overline{Y}_{k-1}(\overline{a}) = 0\right)\right) = \theta_0(k) + \theta_1\left(\frac{\sum_{t=1}^{6} I(a_t = 2)}{6}\right) + \theta_2\left(\frac{\sum_{t=1}^{6} I(a_t = 3)}{6}\right) + \theta_3\left(\frac{\sum_{t=7}^{k-1} I(a_t = 2)}{k-7}\right) + \theta_4\left(\frac{\sum_{t=7}^{k-1} I(a_t = 3)}{k-7}\right),$$

Example 1: Neighborhood effect (Wodtke 2013; Wodtke, Harding, and Elwert, 2011)

Table 4 Effects of neighborhood poverty on the risk of adolescent parenthood, Panel Study of Income Dynamics[a]

	Blacks (person-years = 15,420)				Nonblacks (person-years = 21,548)			
	Regression-Adjusted		IPT-Weighted		Regression-Adjusted		IPT-Weighted	
Model	LOR	SE	LOR	SE	LOR	SE	LOR	SE
Model 1								
Cumulative exposure								
Low-poverty neighborhood	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Moderate-poverty neighborhood	0.467	0.300	0.569	0.320†	0.383	0.269	0.460	0.265†
High-poverty neighborhood	0.455	0.261†	0.601	0.266*	0.571	0.306†	0.829	0.297**
Model 2								
Cumulative exposure, childhood								
Low-poverty neighborhood	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Moderate-poverty neighborhood	0.167	0.329	0.103	0.355	0.129	0.345	0.249	0.364
High-poverty neighborhood	0.152	0.319	0.212	0.339	−0.258	0.449	−0.138	0.478
Cumulative exposure, adolescence								
Low-poverty neighborhood	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Moderate-poverty neighborhood	0.279	0.260	0.422	0.283	0.224	0.263	0.189	0.266
High-poverty neighborhood	0.284	0.269	0.365	0.293	0.670	0.337*	0.790	0.341*
Model 3								
Point exposure (age 11)								
Low-poverty neighborhood	Ref.	Ref.	—	—	Ref.	Ref.	—	—
Moderate-poverty neighborhood	0.213	0.186	—	—	0.167	0.193	—	—
High-poverty neighborhood	0.191	0.178	—	—	0.549	0.218*	—	—

[a]Log odds ratios (LOR) and standard errors (SE) are combined estimates from five multiple imputation data sets.

Example 2: Parental Incarceration and Children's Academic Achievement (Fox, Moore, and Song 2022)

- Data: PSID, PSID Child Development Supplement (CDS), and Fragile Families and Child Wellbeing Study
- Outcome (Y): Children's Academic Achievement
- Time-varying exposure (A, ...Ak): parental incarceration
- Time-varying confounders (La .....Lk)
- Baseline characteristics (*C*)

Example 2: Parental Incarceration and Children's Academic Achievement (Fox, Moore, and Song 2022)

Pre-childbirth incarceration variable, A0, and all Am into a post-childbirth measure of parental; incarceration between age 0 and 18 (or the age when the last CDS measure is observed), A1

$$\mathbb{E}(Y_1 \mid A_0, A_1) = \alpha_{00} + \beta_{00} A_0 + A_1 \Big(\beta_{10} + \beta_{11} A_0\Big) \tag{1}$$

More generally, we estimate models with $t = 1$ (childhood) and $t = 2$ (early adulthood). The model relies on the regression framework based on a weighted pseudopopulation in which the time-varying covariates have been properly balanced across treatment and control groups at each time point.

$$\begin{aligned} \mathbb{E}(Y_2 \mid \bar{\mathbf{A}}) = \alpha_{00} + \beta_{00} A_0 + A_1 \Big(\beta_{10} + \beta_{11} A_0\Big) + \\ + A_t \Big(\beta_{10} + \beta_{11} A_0 + \cdots + \beta_{tt} A_{t-1}\Big) \end{aligned} \tag{2}$$

Example 2: Parental Incarceration and Children's Academic Achievement (Fox, Moore, and Song 2022)

Table 3. MSM Estimated Effects of Parental Incarceration on Academic Achievement, Age 0–18

	Black			White		
	LW	PC	AP	LW	PC	AP
Parental incarceration before birth	1.993	-2.709	-2.318	-20.174*	-18.860†	-14.990**
	(2.904)	(2.513)	(2.126)	(8.974)	(10.147)	(5.369)
Parental incarceration during childhood	-0.465	-1.472	1.051	-8.242***	-4.130	-5.299
	(1.830)	(1.476)	(2.343)	(2.132)	(2.933)	(3.679)
Intercept	97.610***	96.740***	95.833***	107.071***	105.349***	107.934***
	(0.299)	(0.300)	(0.275)	(0.272)	(0.267)	(0.257)
Observations	3,322	2,809	3,309	4,281	3,689	4,268

Table 4. MSM Estimated Effects of Incarceration Timing (Pre-natal vs. Childhood) on Academic Achievement, Age 0–18

	White			Black			Hispanic		
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)
	PC	AP	LW	PC	AP	LW	PC	AP	LW
Par. Inc. 0-3 years before birth	-6.982**	-3.657	-4.310	-14.268	-3.390	-3.311	0.790	-2.460	3.077
	(2.466)	(3.291)	(3.861)	(11.716)	(2.601)	(5.060)	(2.598)	(4.681)	(5.697)
Par. Inc. 0-9 years after birth	-10.447**	-8.802*	-5.596*	-4.415*	-4.940†	-6.533	0.419	-0.483	9.593
	(3.327)	(4.008)	(2.540)	(2.227)	(2.563)	(4.958)	(4.327)	(4.046)	(7.702)
Timing. 0-3 years before birth	2.716	0.257	-8.083†	3.200	0.810	7.070	0.641	1.031	-7.084†
	(2.526)	(2.977)	(4.395)	(4.783)	(1.465)	(4.796)	(1.066)	(2.436)	(4.274)
Timing: 0-9 years after birth	-0.540	-1.131†	-0.821	0.823*	0.479	0.173	0.204	1.494*	-0.877
	(0.668)	(0.630)	(1.078)	(0.355)	(0.615)	(0.481)	(0.574)	(0.629)	(1.119)
Intercept	102.900***	108.790***	105.860***	93.541***	100.080***	106.230***	92.235***	99.187***	96.828***
	(1.249)	(1.199)	(1.765)	(1.673)	(1.794)	(4.765)	(1.735)	(2.087)	(2.388)
Observations	401	402	259	630	633	457	389	395	254

Source: Fragile Families and Child Wellbeing Study 1998–2017.
Note: Parental incarceration before and after birth refer to dummy variables that show the effect of average change in scores associated with a parental incarceration spell during the specified age-based time interval. Incarceration timing before and after birth refer to continuous variables that show whether the timing within the interval (early vs. later) influences the effect significantly. Estimates show the effect associated with having an incarceration spell one year later in the interval. Standard errors are included in parentheses. Other covariates in the model are illustrated in Table 2. The OLS results are presented in Appendix Tables F, G, and H.
$^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$ (two-sided tests).

Example 3: Censoring weight

```
#CD4 count has an effect both on dropout and mortality, which causes informative censoring.

#Use inverse probability of censoring weighting to correct for effect of CD4 on dropout.

#Use Cox proportional hazards model for dropout.

censorm <- ipwtm(
  exposure = dropout,   family = "survival",
  numerator = ~ sex + age,
  denominator = ~ sex + age + cd4.sqrt,
  id = patient,  tstart = tstart,  timevar = fuptime,  type = "first", data = haartdat)
```

tstart: numerical vector, representing the starting time of follow-up intervals, using the counting process notation.

R Codes and data

R codes:
https://www.dropbox.com/s/w7o46d4dtvwf8y4/ipw_demo_Nov18.Rmd?dl=0
Data:
https://www.dropbox.com/s/adw9i7n7a1b5q0w/mydata_example.csv?dl=0